

CLAIMS

What is claimed is:

1. An image reading apparatus for reading an image which contains character information, the apparatus comprising:

labeling process unit to group a continuous black pixel area forming characters contained in a read black and white monochrome image of two gray levels, and extracting group bounding rectangle information about a grouped continuous black pixel area;

overlap integrating process unit to determine overlap between grouped group bounding rectangles, and integrating overlapping group bounding rectangles; and

language determining process unit to obtain a ratio of the number of group bounding rectangles integrated in an overlap integrating process to the number of group bounding rectangles before the overlap integrating process, and determining a language from a characteristic of the overlap integration ratio.

2. The image reading apparatus according to claim 1, further comprising:

row extracting process unit to extract row rectangle information from position information about a group bounding rectangle of the continuous black pixel area extracted and grouped by the labeling process unit when a document includes graphics and pictures,

wherein the overlap integrating process and the language determining process are performed on a group bounding rectangle contained in a row rectangle extracted by the row extracting process unit.

3. The image reading apparatus according to claim 2, further comprising:
binarizing process unit to binarize multi-valued image data when an image read by an image input device is multi-valued image such as a color image, a multilevel gray scale image, etc.
4. The image reading apparatus according to claim 3, further comprising:
statistical determination process unit to perform a language determining process of determining a language from the overlap integration ratio on a plurality of rows contained in an original, and determining in a statistical process a language determined as a language of characters contained in most rows as a language of characters contained in the original.
5. The image reading apparatus according to claim 2, further comprising:
statistical determination process unit to perform a language determining process of determining a language from the overlap integration ratio on a plurality of rows contained in an original, and determining in a statistical process a language determined as a language of characters contained in most rows as a language of characters contained in the original.
6. The image reading apparatus according to claim 1, further comprising:
binarizing process unit to binarize multi-valued image data when an image read by an image input device is multi-valued image such as a color image, a multilevel gray scale image, etc.

7. The image reading apparatus according to claim 6, further comprising:
statistical determination process unit to perform a language determining
process of determining a language from the overlap integration ratio on a plurality
of rows contained in an original, and determining in a statistical process a
language determined as a language of characters contained in most rows as a
language of characters contained in the original.